Comparison of Quantitative and Visual Detection of New Focal Bone Lesions

W. Robert Pitt* and Peter F. Sharp

Department of Bio-Medical Physics and Bio-Engineering, University of Aberdeen, Foresterhill, Aberdeen, Scotland

A comparison is made between the sensitivity of detection of focal abnormalities in radionuclide bone images by visual examination and by using simple quantification. The quantitative method calculates the ratio of radiopharmaceutical uptake in a region of interest drawn around a lesion to that in an area of normal bone. Quantification is found to be far more sensitive than visual examination in detecting focal metastases. The use of "baseline" images improved the precision of quantitation of rib lesions, but appeared not to alter the sensitivity of visual detection. This method of quantification is therefore limited more by the inability of observers to notice suspicious areas to which it should be applied than by inaccuracies inherent in the method itself. Further work should concentrate more on image enhancement than on improving quantitative techniques.

J Nucl Med 26:230–236, 1985

Radionuclide imaging of the skeleton ("bone image") is the most frequently performed investigation in nuclear medicine. Most skeletal studies are used to detect or stage progression of metastatic disease from primary breast tumors. Usually the images are assessed visually by trained observers, although modern data-processing systems offer facilities for several types of quantitative analysis. This paper describes experiments to compare the relative sensitivity of a simple quantitative technique against visual assessment for the detection of focal areas of increased count density.

The clinical value of the work to be described rests on the assumption that a useful indicator of the progression of metastatic bone disease in patients with breast cancer is the appearance of new lesions, rather than in any change of uptake in established metastases (*1*). Although this opinion is open to question, discussion of such controversy is outside the scope of this report which is concerned simply with improving the detection of low-contrast foci; the acknowledged nonspecific nature of such lesions makes it necessary for each one to be assessed with full consideration of other clinical indicators.

Received Oct. 12, 1983; revision accepted Nov. 1, 1984.

For reprints contact: P. F. Sharp, BSc, PhD, Dept. of Bio-Medical Physics and Bio-Engineering, University of Aberdeen, Foresterhill, Aberdeen AB9 2ZD, Scotland.

* Present address: International General Electric Medical Systems, Colney St., Radlett, St. Albans, Hertfordshire, AL2 2ER, England.

Most quantitative methods applied to bone images involve comparing mean count densities in small, operator-defined regions of interest (ROIs) in the image. One ROI is marked around the area of the suspected metastasis, and the other over a similar area of either soft tissue or bone. The use of soft tissue as a reference has been found unreliable due to variations in soft-tissue uptake of bone imaging agents (*2*). Instead, the target-to-bone ratio (t/b) is generally preferred, in which an area of bone that is assumed to be normal is used as the reference. In cases of disseminated bone disease it is not possible to assume any bone is normal, but as metastases from breast cancer are usually focal, the technique is justifiable in this case. In this study quantification will be carried out using t/b, a technique similar to that used by several others (*3,4*).

The effectiveness of visual assessment of an image will be measured in terms of the minimum contrast needed to detect a lesion, using the method of constant stimulus (*5*).

Baseline images, i.e., images collected as near as possible to the time of diagnosis of the primary tumor, have been used to assist in the interpretation of bone images, but their value is controversial (*6–10*).

If a baseline image is available then it can be compared with a later image to reduce the effect of "anatomical noise"—i.e., uncertainty due to variations between the normal uptake patterns of different patients

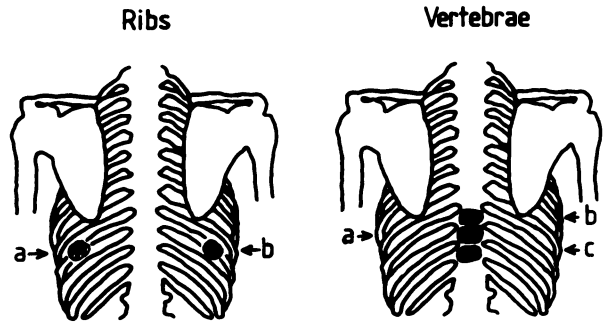

FIGURE 1
Determination of t/b values, R. For single images, taking c_a to be number of counts in ROI "a" and n_a as number of pixels in that ROI, then

$$R = \frac{c_a/n_a}{c_b/n_b}$$

in ribs and

$$R = \frac{c_a/n_a}{\frac{1}{2}\left\{\frac{c_b}{n_b} + \frac{c_c}{n_c}\right\}}$$

in spine. With baseline images, $R' = R_2/R_1$, where R_1 was obtained from baseline image, and R_2 from later one

from such benign causes as scoliosis or occupational stress (*10*).

Visual assessment can be aided by display of the baseline and later image side by side, while in the quantitative approach values of t/b calculated from the same anatomical sites in the two images can be compared.

This paper will investigate whether the use of the baseline images improves significantly the detection of lesions.

METHODS

Upper posterior-view skeletal images, acquired at least 2 hr after patients had been injected with technetium-99m- (^{99m}Tc) labeled methylene diphosphonate (MDP), were used in all these experiments.

Quantitative analysis of single images

Fifteen normal bone images were used in this experiment. The images were reported normal in 1981 when they were made, and the subsequent medical history of each patient was followed up. None of them showed clinical evidence of bone disease in the subsequent 2 yr, so the images were all considered normal.

The images were digitized in 128 × 128 matrices, and 50 irregular ROIs were marked over sections of the ribs. The total number of counts and number of elements within each ROI were recorded. The same information was obtained from similarly shaped ROIs marked on the contralateral ribs. For each pair of ROIs, the ratio of mean number of counts per element, R, was calculated as:

$$R = \frac{c_a/n_a}{c_b/n_b} \qquad (1)$$

where c_a is the number of counts in ROI "a" and n_a the number of elements in the ROI. ROI "a" and "b" are as shown in Fig. 1.

A similar type of analysis was performed on 20 vertebrae. Since no contralateral area existed, the mean number of counts per element in an ROI over the vertebra of interest was compared with the average from similar ROIs drawn over vertebrae immediately above and below it. Again the ratio of these two values, R, was calculated as:

$$R = \frac{c_a/n_a}{\frac{1}{2}\left\{\frac{c_b}{n_b} + \frac{c_c}{n_c}\right\}}, \qquad (2)$$

where ROIs a, b, and c are as shown in Fig. 1.

This procedure was performed independently by two experienced markers.

Visual assessment of single images

Before quantitative analysis can be applied to an image, lesion contrast must be high enough for an observer to suspect that an area is abnormal. This critical contrast was determined as follows.

Thirty images were modified by adding clinically realistic artificial "lesions" produced by a computer simulation. The method is described in detail elsewhere (*11*) and involved first marking a realistic outline of the lesion on the image and then adding Poisson-distributed random numbers to the area to give a focal "lesion" of the required contrast. Lesion contrast was defined as the ratio of the increase in count density to the count density that would be found in the same area if the image were normal. This is different from t/b which compares count densities in different (contralateral or adjacent) parts of an image. Small lesions (foci of increased count density) were added to the ribs, and large ones, usually encompassing whole vertebral bodies, to the spine. The contrasts of the small abnormalities increased in six equal steps from 20 to 95%, and 30 foci of each contrast were created. The contrasts of the large abnormalities increased in five equal steps from 10 to 50%, and 18 of each contrast were added.

The images, displayed on a 128 × 128 matrix, were presented to four clinically experienced observers. Each image was shown simultaneously on two monochrome TV monitors, arranged so that the visual angles of the images were equivalent to viewing a square image of side length 32 cm from 4 and 8 m at the same time. This has been shown to be the best arrangement for detecting focal lesions in bone images (*11*). The TV monitors were interfaced to a microprocessor that permitted interactive image manipulation (*12*), observers being allowed to scale the apparent image count density and alter the brightness and contrast of the TV display. Using a trackerball controlling a cursor on the screens, each observer marked the center of any area in an image that he considered was either definitely abnormal, or alter-

natively, suspicious enough to justify the use of quantitative analysis. He then pressed one key of a computer terminal if he was certain a lesion was present, and another if he was just suspicious. A response was recorded as true-positive if the cursor was not more than one pixel outside the boundary of the lesion.

Quantitative analysis using the baseline image

Some previously published work concerning quantitation of repeat images (*1*) has involved images collected many months apart, as would be done in routine clinical practice. For experimental purposes, however, these data are not satisfactory since the real underlying changes in the clinical state of the patient between the images cannot be known. Thus if a second image shows a definite metastasis that was not seen on the first image, it cannot be determined whether the lesion was present at the time of the first image and overlooked (a false-negative response), or whether it was absent at that time but developed between images (a true-negative response). One solution would be to inject and image a patient on two successive days when the physiology would have changed little, but other variables (such as injection quality, patient position, camera response) would have altered realistically. As this would be unethical, a practical compromise was adopted.

For several months a second upper posterior image was acquired from every patient who attended this center for bone imaging. In the few minutes between the two acquisitions the patient walked about and the camera was moved out of position and back again, so producing realistic variations in patient positioning between repeat images. No account was taken of variations in the quality of the radiopharmaceutical because the standard of labeling of MDP at this center is very high. Daily variations in camera performance are also absent.

Thirty pairs of bone images were selected, all of which had been passed as normal by clinically experienced staff on two separate occasions. These pairs of images were used in experiments that were concerned only with differences between each pair and so did not require a 2-yr follow-up.

Numerical analysis was performed as before, but this time ROIs were marked on both images of each pair. By using photographs showing the ROIs superimposed on the first image, it was possible to mark corresponding areas on the second image with confidence. This time, the ratio R′, was calculated as:

$$R' = R_2/R_1,$$

where R_1 is the ratio R [Eq. (1) or (2)] taken from the baseline image and R_2 the ratio from the equivalent area on the later image. In this way it did not matter if any of the patients had undetected low-level bone abnormalities, since all features common to both images cancelled out. Two markers again produced 50 ratios from ribs, and 20 from vertebrae.

Visual assessment using the baseline image

Images modified in the manner described above were also used to investigate whether lesions of lower contrast could be detected when baseline images were available for comparison. As with the investigation of baseline images in quantitation just described, all the modified images were originally one of a pair collected within minutes of each other. In this experiment each modified image was displayed alongside the second, unmodified one of the pair which was used as the baseline. An example of such a pair is shown in Fig. 2. Therefore, although the images were not identical (due to changes in camera and patient positions), the only clinically relevant differences between them were the artificial abnor-

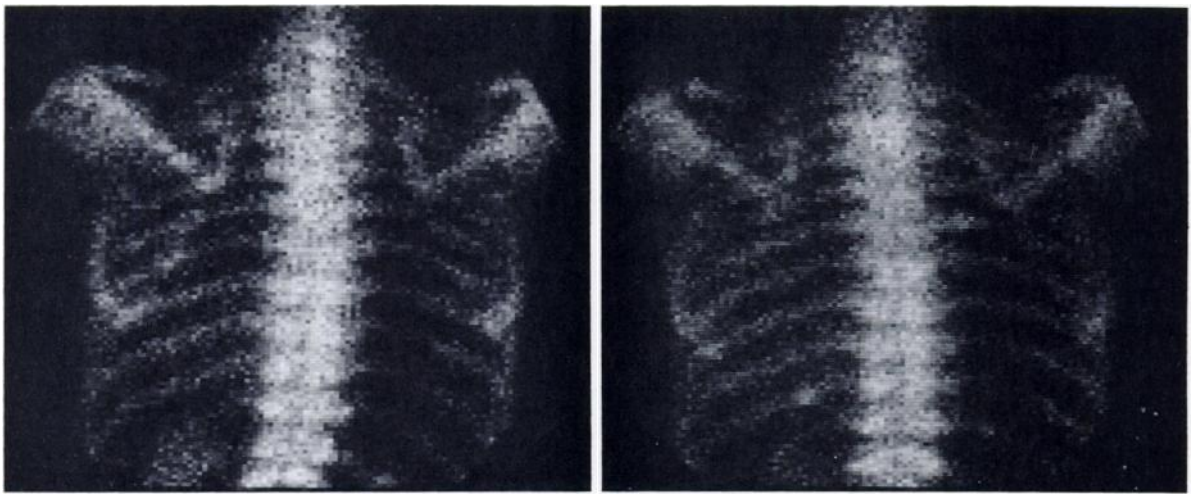

FIGURE 2
Typical pair of images of same patient collected few minutes apart. Unmodified image acted as baseline image in visual experiments, while modified one has had three "lesions" added to spine and six added to ribs

TABLE 1
Contrasts Required (As Percentages) for Lesions to be Detected as Abnormal by Quantitation

Position of lesions	Marker 1		Marker 2	
	Single images	With baseline images	Single images	With baseline images
Ribs	51	30	47	33
Spine	11	17	17	14

malities that had been added. Observers were asked to use the baseline image in assessing the abnormal one. Apart from this change, observers followed the procedure already described.

RESULTS

Quantitative analysis of single images

To find the limiting values for R in normal images, i.e., the values above or below which a result would represent an abnormal ratio, the 2.5 and 97.5% percentiles were calculated from the samples (*13*). The results from the ribs produced an upper limit on R of 1.22 for Marker 1 and a lower limit of 0.81. For Marker 2, the upper limit was 1.16, and the lower limit 0.78.

Consider next the measurements on the spine. The upper limit of R for Marker 1 was 1.06, and the lower limit 0.96. For Marker 2, the corresponding values were 1.06 and 0.91.

The most difficult case to which this method may be applied is when a genuine lesion is present in an area of an image that would, in the absence of the lesion, be at the lower limit of the normal range of R. Under these circumstances, the contrast of the abnormality would have to be high enough for the measured ratio R to exceed the upper limit. For rib lesions, this contrast would be 51% (1.22/0.81-1) for Marker 1, and 47% (1.16/0.78-1) for Marker 2. For vertebral measurements the contrast would need to exceed 11% (1.06/0.96-1) for Marker 1 and 17% (1.06/0.91-1) for Marker 2. These results are summarized in Table 1. This is the worst case, so these contrasts are the maximum needed for the normal limits to be exceeded. Any lesions with contrasts higher than these values, and indeed many that were lower, would be shown to be abnormal (at the 95% level) by this simple quantitative method.

Visual assessment of single images

The results from all four observers were pooled and expressed as percentage true-positive response rates. These results, classified by the contrast of the abnormalities and the degree of certainty with which they were seen, are shown in Table 2 for the small rib lesions, and in Table 3 for the large vertebral ones.

Visual response curves, discussed in detail elsewhere (*5*), were calculated from the results of viewing the large and small abnormalities at both degrees of certainty. These sigmoid curves may be represented by cumulated normal distributions, and a weighted best-fit line was calculated for each set of results using the maximum likelihood methods of probit analysis (*14*). The chi-squared test applied to the curves showed that the fits were adequate ($p > 0.05$). The detection contrast, defined as the contrast which elicits from an average observer a 50% true-positive visual response, is equivalent to the visual threshold adopted by the average observer (*5*), and this was calculated from the best fit curves. The curves for both degrees of certainty are shown in Fig. 3. The detection contrast for foci in the ribs was 97 ± 5% for definite detection, and 61 ± 3% for suspicion. [The errors are 95% confidence limits (*14*).] For vertebral bodies, the detection contrast was 40 ± 2% for certain detection, and 26 ± 2% for suspicion. On average, therefore, half of the lesions with these contrasts in a bone image would be thought by observers to be definitely or possibly abnormal, respectively.

Quantitative analysis using the baseline image

The 97.5% percentile values of R′ taken from the ribs were 1.14 and 1.16 for Markers 1 and 2, respectively. The corresponding 2.5% percentile values were 0.88 and 0.87. The upper limits from the spine were 1.09 and 1.06 for Markers 1 and 2, respectively, and the lower limits were 0.93 for both markers.

TABLE 2
True-Positive Visual Response Rates (As Percentages) from Small Rib Lesions

Contrast (%)	From single images		With baseline images	
	Suspicious	Certain	Suspicious	Certain
20	1	—	2	—
35	14	2	17	3
50	39	3	38	4
65	62	17	58	25
80	75	23	76	31
95	88	48	89	55

TABLE 3
True-Positive Visual Response Rates (As Percentages) From Large Vertebral Lesions

Contrast (%)	From single images		With baseline images	
	Suspicious	Certain	Suspicious	Certain
10	11	3	18	3
20	35	7	32	8
30	61	21	61	26
40	94	57	97	64
50	91	75	97	83

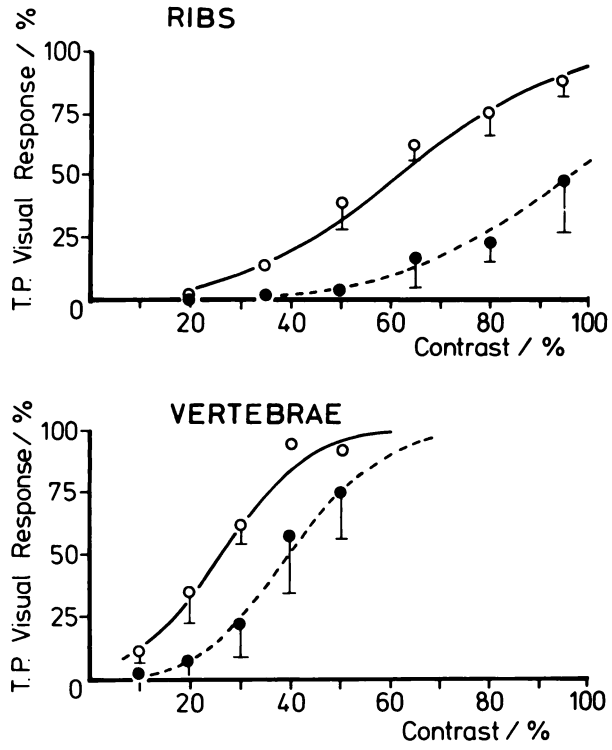

FIGURE 3
Visual response curves for observers' pooled results from single images (Tables 2 and 3). Solid circles and dashed lines are results when observers were certain of presence of abnormality, and open circles and solid lines are results when they were only suspicious of its presence. Error bar shows variability between observers (1 s.d.)

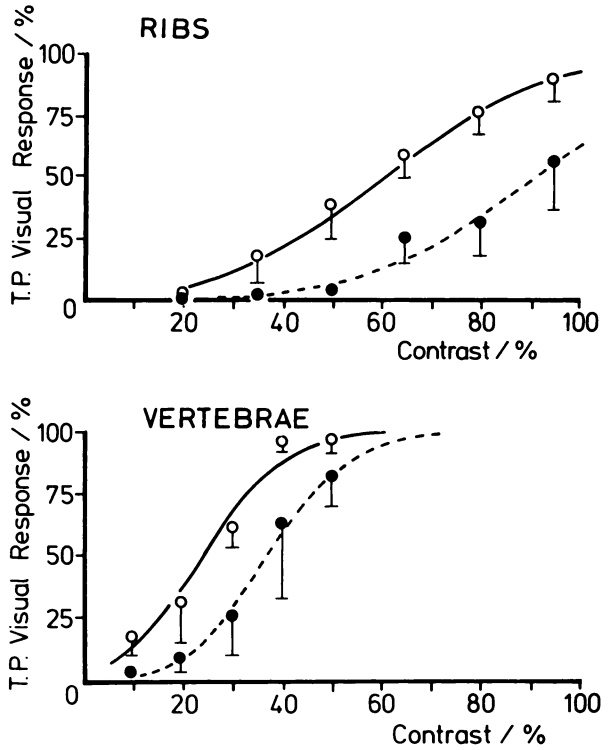

FIGURE 4
Visual response curves for observers' pooled results from using baseline images (Tables 2 and 3). Symbols are explained in Fig. 3

From the argument developed above, the maximum abnormality contrast required for a rib lesion to exceed 95% probability of detection is 30% (1.14/0.88-1) for Marker 1 and 33% (1.16/0.87-1) for Marker 2. Corresponding values for vertebral lesions are 17% (1.09/0.93-1) and 14% (1.06/0.93-1). These results are summarized in Table 1.

Visual assessment using the baseline image

The pooled results, again classified by abnormality contrast and each observer's degree of certainty are also shown in Table 2 for the small rib lesions and in Table 3 for the large vertebral ones.

Visual response curves for the results were calculated as before, and are shown in Fig. 4. The detection contrasts for lesions in the ribs were 91 ± 5% for definite detection, and 61 ± 3% for suspicion. For vertebral bodies these values were 37 ± 2% and 24 ± 2%.

DISCUSSION

Comparison of quantitative and visual methods for detecting focal lesions

The quantitative experiment applied to single images showed that all rib lesions with contrasts greater than about 50%, and all spine lesions greater than 11% would be identified as abnormal (with 95% certainty) by the simple method described (Table 1). On average, however, only about one in three rib lesions and one in five vertebral ones at these contrasts would even be regarded as suspicious when viewed by clinically experienced observers, and far fewer would be classed as definitely abnormal. Thus it appears that, using single images, the simple quantitative method would be of value in confirming or refuting the presence of lesions in areas that were seen to have only slightly elevated count densities.

Effect of baseline images on quantitative results

The relative precision of t/b measurements made with and without baseline images may be judged from Table 1. This shows that both markers were able to detect rib lesions of lower contrast when baseline images were used. This effect does not seem to be consistent for vertebral lesions.

The statistical validity of any improvement may be determined by comparing the spread of the t/b values in each case (*15,16*). This will show whether the variability in R measured from single images differs from that of R′ obtained when baseline images were introduced. The images used for the two experiments were

completely different, so the results are uncorrelated and a more precise analysis cannot be performed.

If we consider ROIs drawn in the ribs, the spread in the values of R was significantly greater than those of R′ for both markers ($p < 0.05$). A comparison of the spread of the vertebral measurements, however, showed no significant difference at the 95% level.

The use of baseline images therefore significantly improved the precision of t/b measurements on the ribs, but not on the larger areas in the spine.

Effect of baseline images on visual results

If detection contrast is taken as the criterion for comparing observer performance (*5*), baseline images produced no improvement. For rib lesions the "certainty" threshold altered from 97 ± 5% to 91 ± 5%, and the threshold of suspicion remained the same at 61 ± 3%. For vertebral bodies the corresponding changes were from 40 ± 2% to 37 ± 2% and from 26 ± 2% to 24 ± 2%.

The use of a single threshold to describe the process of visual detection has been criticized by advocates of signal-detection theory (*17*). Experiments that impose a single threshold have been used here because this method provides a comparison of visual detection with quantitation. We hope to report experiments employing receiver operating characteristic analysis to investigate more fully the effect of baseline images on visual detection. Such an investigation, however, is outside the scope of this paper.

CONCLUSIONS

Although only a small number of observers participated in this work, it appears that they were far less effective than even a simple quantitative method in detecting low-contrast focal lesions. On average, for every three low-contrast lesions in the ribs (or five in the spine) that would definitely be classified as abnormal by quantification, only one would even be considered suspicious by an observer.

The use of baseline images significantly improved the precision of t/b measurements of rib lesions, but not of those in the spine. No improvement in visual detection was apparent, although further work is in progress. Thus baseline images increase the difference in sensitivity between quantitative and visual detection.

Various improvements to quantitative methods have been suggested, including the use of an external source to calibrate the camera (*18*), a transmission image of each patient (similar to a radiograph) to determine attenuation factors (*19*), and careful weighing, dilution and imaging of a sample of the injected radiopharmaceutical to measure accurately the activity administered to each patient (*20*). Although these refinements are important in some circumstances, such as when diffuse bone disease is suspected, the results reported here show that the "rough and ready" method of ROI analysis is more accurate and useful in detecting focal lesions than might be expected. Furthermore, since no ROI method can be applied unless an observer is suspicious of part of an image, it appears more important to concentrate future attention on ways of enhancing the visual appearance of images so that observers can make full use of simple quantitative methods, rather than on more complicated and time-consuming improvements to the methods themselves.

In conclusion, it appears that quantitation is much more sensitive than visual detection of lesions particularly if baseline images are used. Low-contrast lesions may pass undetected because observers do not see them and so do not apply quantitative analysis, rather than because the quantitative analysis is not sufficiently precise. Improvements in image processing are therefore needed before even the crudest methods of quantitative analysis can achieve their full potential in this area of diagnosis.

ACKNOWLEDGMENTS

The programs for adding the artificial lesions to the bone images were largely the work of Mr. R. B. Chesser. The authors are also grateful to the following additional observers: Drs. H. Gemmell, M. Y. Norton and Mrs. J. Davidson, and to Professor J. R. Mallard for his encouragement. At the time this work was carried out, one of the authors (WRP) was in receipt of a Research Studentship from the Medical Research Council.

REFERENCES

1. Condon BR, Buchanan R, Garvie NW, et al: Assessment of prognosis of secondary bone lesions following cancer of the breast or prostate using serial radionuclide imaging. *Br J Radiol* 54:18–23, 1981
2. Pfeifer JP, Buell U, Pfeifer H: Quantitative assessment of ^{99m}Tc-MDP scans in the investigation of diffuse alterations in bone. *Eur J Nucl Med* 4:407–412, 1979
3. Sneppen O, Heerfordt J, Dissing I, et al: Numerical assessment of bone scintigraphy in primary bone tumors and tumor-like conditions. *J Bone Joint Surg (Am)* 60A:966–969, 1978
4. Büll U, Keyl W, Meister P, et al: Wertigkeit der "region of interest"—Technik in der skelettszintigraphischen diagnostik primärer knochentumoren. *Radiologe* 21: 46–51, 1981
5. Sharp PF, Chesser RB, Smith DB: Assessment of scintigram quality. *IEE Proc* 127(Pt. E):185–192, 1980
6. Clark DG, Painter RW, Sziklas JJ: Indications for bone scans in preoperative evaluation of breast cancer. *Am J Surg* 135:667–670, 1978
7. McNeil BJ: Rationale for the use of bone scans in selected metastatic and primary bone tumors. *Semin Nucl Med* 8:336–345, 1978
8. Calle R, Gongora R, Vilcoq JR, et al: Cancers du sein: Diagnostic des metastases infra-cliniques osseuses.

Valeur previsionnelle de la scintigraphie. *Bull Cancer (Paris)* 67:313–317, 1980
9. Coker DD, Lambrecht RW, Kehn BD: The value of initial and follow-up bone scans in patients with operable breast cancer. *Milit Med* 145:492–494, 1980
10. Muroff LR: Optimizing the performance and interpretation of bone scans. *Clin Nucl Med Suppl* 6:P68–P76, 1981
11. Pitt WR, Sharp PF, Chesser RB, et al: Radionuclide image minification can compensate for coarse digitization: Concise communication. *J Nucl Med* 24:1046–1054, 1983
12. Smith DB, Undrill PE, Bolton MP: Optimising the performance of medical diagnostic displays by microprocessor assistance. In *Changes in Health Care Instrumentation due to Microprocessor Technology*, Pinciroli F, Anderson J, eds. Amsterdam, North Holland Publishing Company, 1981, pp 119–126
13. Diem K, Lentner C, eds: *Scientific Tables*, Basle, Ciba-Geigy Ltd., 1970, pp 160–161
14. Finney DJ: *Probit Analysis, Second Edition*, Cambridge, Cambridge University Press, 1952
15. Wilks SS: Statistical prediction with special reference to the problem of tolerance limits. *Ann Math Stat* 13: 400–409, 1942
16. Rosenbaum S: Tables for a nonparametric test of dispersion. *Ann Math Stat* 24:663–668
17. Tanner WP, Swets JA: A decision making theory of visual detection. *Psychol Rev* 61:401–409, 1954
18. Lurye DR, Castronovo FP, Potsaid MS: An improved method for quantitative bone scanning. *J Nucl Med* 18:1069–1073, 1977
19. Graham LS, Neil R: In vivo quantitation of radioactivity using the anger camera. *Radiology* 112:441–442, 1974
20. Hardy JG, Kulatilake AE, Wastie ML: An index for monitoring bone metastases from carcinoma of the prostate. *Br J Radiol* 53:869–873, 1980